The Forecasting Performance of Seasonal and Nonlinear Models

Author
Houda Ben Hadj Boubaker
The Higher Institute of Management of Tunisia, Tunisia

Key words: Seasonality, Nonlinearity, Interval forecasts, Density forecasts, Forecast evaluation tests.

JEL Classification: C22, C52, C53

Abstract

In this paper, we compare the forecasting performance of seasonal and non linear autoregressive models in terms of point, interval, and density forecasts for the growth rates of the Tunisian industrial production, for the period 1976:1- 2006:2. Our results suggest that the point forecasts generated by the linear models perform better than those provided by the nonlinear models at all horizons. By contrast, the analysis of interval and density forecasts at horizons of one and three quarters provide an evident support for the nonlinear models, this result is in line with the literature. Thus, our findings assess the usefulness of nonlinear models to investigate the dynamic behavior of economic systems and to produce accurate forecasts.

Introduction

Macroeconomic variables such as unemployment, industrial production, investment and exchange rate display features that seem to require nonlinear time series models. These features include asymmetric behavior of the series over the business cycle, regime switching and volatility which can not be detected by linear models, so it is not surprising that nonlinear models gained more attention and became a useful tool to forecast economic time series. Examples of these models are the markov switching model developed in Hamilton (1989), the threshold autoregression models proposed by Tong (1990) and the smooth transition autoregression models advocated in Teräsvirta and Anderson (1992). In addition seasonal fluctuation is a dominant feature of monthly or quarterly macroeconomic time series. Hence seasonality should be considered in the forecasting exercise. One of the ultimate uses of these models is to reduce uncertainty about the possible future projections. Thus, it is imperative that we have better forecasts. To this end, discriminating among the competing models is required. In the literature the forecasting performance focused mainly on point forecasts see Clements and Krolzig (1998), Stock and Watson (1998) and Marcellino (2004). Recently evaluating interval and density forecasts received increasing attention. As provided by De Goojer and Kumar (1992) that the point forecasts performance of non linear models are no worse or better than linear models. Many other studies proved the nonlinear models superiority, for example, Medeiros et al.(2001) comparison forecast's conclusion are in favor of the nonlinear models using Smooth Transitions and Neural Networks and model applied on several monthly exchange rates time series. Marcellino (2004) compares the forecasting performance of many fitted nonlinear and time-varying models of the EMU macroeconomic variables, with linear models. He argued that the nonlinear model perform well. Siliverstovs and van Dijk (2003) state that the linear models perform well when they compare point forecasts but they fail to render more accurate interval and density forecasts than nonlinear models.

The remainder of this paper is as follows. In section 2, we describe the seven linear and non linear models under evaluation. The different criteria to evaluate the performance of these models are discussed in section 3. In section 4, we discuss the data and the empirical results and finally, section 5 concludes.

Models presentation

We analyze the quarterly growth rates of the industrial production index after transforming the data by taking logarithm. We also assume the presence of unit root at the zero frequency and we detrend the series by first differencing. The specification of the first three linear models consist in selecting the lag operator p as the order that minimizes the Bayesian Information Criteria (BIC). The residual autocorrelation is examined using the Lagrange Multiplier (LM) test at the 5% significance level. The nonlinear models specification will be given for each (see Franses and van Dijk (2005) for more details).

A linear model with stable deterministic seasonality

The first model, assumes the constant of the seasonal components. The use of seasonal dummies is motivated by their simplicity and their ability to capture a constant stable seasonal component. This model, labeled as AR in the following, reads as:

$$\phi(L)\Delta_1 y_t = \delta_1 D_{1,t} + \delta_2 D_{2,t} + \delta_3 D_{3,t} + \delta_4 D_{4,t} + \varepsilon_t .....(1)$$

Where $\phi(L)$ is a polynomial of order p in the lag operator L with all roots outside the unit circle, Δ_1 is the differencing operator defined as $\Delta_k y_t = y_t - y_{t-k}$ for all integer $k \neq 0$, the seasonal dummies $D_{s,t} = 1,......,4$, defined as $D_{s,t} = 1$ if time t corresponds to season s and $D_{s,t} = 0$ otherwise and $\varepsilon_t \to NID(0,\sigma^2)$.

A linear model with unit roots

The seasonal unit roots model has became somewhat popular to model changing seasonal pattern, it corresponds to the presence of stochastic trends at the seasonal frequencies. Tests for unit roots were first proposed by Fuller (1976) and the extension to the autoregressive presentation was first analyzed by Hylleberg, Engel, Granger and Yoo (1990) [HEGY]. The SUR model can be defined as follows:

$$\phi(L)\Delta_4 y_t = \mu + \varepsilon_t .....(2)$$

Where $\phi(L)$ and ε_t are as defined before, μ is the slope parameter, $\Delta_4 = (1-L)(1+L)(1-iL)(1+iL)$ with $i_2 = -1$, model (2) permits to the seasonal pattern to evolve over time and subsequently the seasonal unit roots leads to a better modeling of the economic policies.

A linear model with smoothly changing deterministic seasonality

Technological change and changes in institutions and habits may cause changes in the seasonal pattern (van Dijk,Strikholm and Teräsvirta 2003), such changes are steady and continuous and need to be modeled using a smooth changing deterministic time varying seasonality. Therefore the model used is the time varying deterministic autoregressive (TVD-AR) model:

$$\phi(L)\Delta_1 y_t = \sum_{s=1}^{4} \delta_{1,s} D_{s,t}(1 - G(t;\gamma,c)) + \sum_{s=1}^{4} \delta_{2,s} D_{s,t} G(t;\gamma,c) + \varepsilon_t .....(3)$$

(1)

Where $G(t;\gamma,c)$ is the logistic function

$$G(s_t;\gamma,c) = \frac{1}{1+\exp\{-\gamma(s_t - c)\}}, \quad \gamma \rangle 0.....(4)$$

As s_t increases, the logistic function changes monotonically from 0 to 1, with the change being symmetric around the location parameters c, as $G(c-z;\gamma,c) = 1 - G(c+z;\gamma,c)$ for all z, the logistic function. The slope parameter γ determines the smoothness of the change. As $\gamma \to \infty$, the logistic function $G(s_t;\gamma,c)$ approaches the indicator function $I[s_t \rangle c]$, whereas if $\gamma \to 0$, $G(s_t;\gamma,c) \to 0.5$ for all values of s_t.

A nonlinear model with stable deterministic seasonality

To capture the business cycle asymmetry, we use models which allow for such regime-switching behavior. We consider a STAR model (Teräsvirta 1994), but with deterministic seasonal patterns for the first differenced time series. The model is:

$$\Delta_1 y_t = \delta_1 + D^*_{1,t} + \delta_2 D^*_{2,t} + \delta_3 D^*_{3,t} + (\mu_1 + \phi_{1,1}\Delta_1 y_{t-1} + ...+ \phi_{1,p}\Delta_1 y_{t-p})(1 - G(\Delta_4 y_{t-d};\gamma,c)) + (\mu_2 + \phi_{2,1}\Delta_1 y_{t-1} + ...+ \phi_{2,p}\Delta_1 y_{t-p}) G(\Delta_4 y_{t-d};\gamma,c) + \varepsilon_t .....(5)$$

Where $D^*_{s,t} = D_{4,t}$, $s = 1,2,3$, and $G(\Delta_4 y_{t-d};\gamma,c)$ with $d \rangle 0$ is a logistic function as in (4). The expansions and recessions correspond to the logistic function when it is equal to zero and one, respectively, and the regime-switches are determined by the variable $\Delta_4 y_{t-d}$. As we remark that the model in (5) is the nonlinear version of the AR model given in (1). Thus, to specify the STAR model, the AR model has to be considered and the value of d should be determined using the LM-type test. Then the autoregressive order p in (5) should be re-specified by the means of the BIC and LM-type test of no first-to-fifth order residual autocorrelation at the 5% significance level.

A nonlinear model with smoothly changing deterministic seasonality:

The model considered is labeled the time varying deterministic STAR (TVD-STAR) model which is an extension of the linear model with smoothly changing deterministic seasonality (3):

$$\Delta_1 y_t = \left[\mu_1\left(1-G\left(\Delta_4 y_{t-d};\gamma_2,c_2\right)\right)+\mu_2 G\left(\Delta_4 y_{t-d};\gamma_2,c_2\right)\right]$$
$$\left(1-G(t;\gamma_1,c_1)\right)+\begin{bmatrix}\mu_3\left(1-G\left(\Delta_4 y_{t-d};\gamma_2,c_2\right)\right)\\ +\mu_4 G\left(\Delta_4 y_{t-d};\gamma_2,c_2\right)\end{bmatrix}G(t;\gamma_1,c_1)$$
$$+\sum_{s=1}^{3}\delta_{1,s}D^*_{s,t}\left(1-G(t;\gamma_1,c_1)\right)+\sum_{s=1}^{3}\delta_{2,s}D^*_{s,t}G(t;\gamma_1,c_1)$$
$$+\left(\phi_{1,1}\Delta_1 y_{t-1}+\ldots+\phi_{1,p}\Delta_1 y_{t-p}\right)\left(1-G\left(\Delta_4 y_{t-d};\gamma_2,c_2\right)\right)$$
$$+\left(\phi_{2,1}\Delta_1 y_{t-1}+\ldots+\phi_{2,p}\Delta_1 y_{t-p}\right)G\left(\Delta_4 y_{t-d};\gamma_2,c_2\right)+\varepsilon_t.....(6)$$

Where $D^*_{s,t} = D_{s,t} - D_{4,t}$, $s = 1,2,3$. As in the specification in (3), the seasonal pattern is allowed to vary over time, but it is assumed to be identical in expansions and recessions at all times. The first step to specify the model in (6) is to start by the TVD-AR model in (3), for which the value of d is selected as the one which minimizes the p-value of the LM-type test statistic against the alternative (6) with $\Delta_4 y_{t-d}$ for $d = 1,...., 4$. After fixing the transition variable, the autoregressive order p in (6) will be re-specified by minimizing the BIC subject to an LM tests.

A time-varying smooth transition model

In order to model continuous structural change and smooth transition-type nonlinearity, Lundbergh,Teräsvirta and van Dijk (2003), combine the STAR model and the time varying autoregressive model, the resultant model is given by:

$$\Delta_1 y_t = \left(\left[\mu_1+\sum_{s=1}^{3}\delta_{1,s}D^*_{s,t}+\phi_{1,1}\Delta_1 y_{t-1}+....+\phi_{1,p}\Delta_1 y_{t-p}\right]\left(1-G\left(\Delta_4 y_{t-d};\gamma_2,c_2\right)\right)\right.$$
$$\left.+\left[\mu_2+\sum_{s=1}^{3}\delta_{2,s}D^*_{s,t}+\phi_{2,t}\Delta_1 y_{t-1}+....+\phi_{2,p}\Delta_1 y_{t-p}\right]G\left(\Delta_4 y_{t-d};\gamma_2,c_2\right)\right)$$
$$\left(1-G(t;\gamma_1,c_1)\right)\left(\left[\mu_3+\sum_{s=1}^{3}\delta_{3,s}D^*_{s,t}+\phi_{3,1}\Delta_1 y_{t-1}+....+\phi_{3,p}\Delta_1 y_{t-p}\right]\right.$$
$$\left.\left(1-G\left(\Delta_4 y_{t-d};\gamma_2,c_2\right)\right)\right)+\left[\mu_4+\sum_{s=1}^{3}\delta_{4,s}D_{*s,t}+\phi_{4,1}\Delta_1 y_{t-1}+....+\phi_{4,p}\Delta_1 y_{t-p}\right]$$
$$G\left(\Delta_4 y_{t-d};\gamma_2,c_2\right)G(t;\gamma_1,c_1)$$
$$+\varepsilon_t.....(7)$$

The TV-STAR model (7) will be specified following the same steps of the TVD-STAR model specification discussed before, but we now start from the AR model.

A nonlinear model with seasonal unit roots

The final model labeled SUR-STAR model combines the seasonal unit root model (2) with smooth transition type regime-switching dynamics, that is,

$$\Delta_4 y_t = \left(\mu_1+\phi_{1,1}\Delta_4 y_{t-1}+....+\phi_{1,p}\Delta_4 y_{t-p}\right)$$
$$\left(1-G\left(\Delta_4 y_{t-d};\gamma,c\right)\right)$$
$$+\left(\mu_2+\phi_{2,1}\Delta_4 y_{t-1}+.....+\phi_{2,p}\Delta_4 y_{t-p}\right)$$
$$G\left(\Delta_4 y_{t-d};\gamma,c\right)+\varepsilon_t.....(8)$$

The delay parameter is obtained by testing the SUR model against the SUR-STAR model. Then, we have to re-specify the model using the same rule to select the lag order p (BIC and the LM type test).

Forecast evaluation

In this section, we employ the existent criteria to evaluate the competing models in terms of point, interval and density forecasts.

Point forecasts

Point forecast evaluation is still receiving attention, Harvey, Leybourne and Newbold (1997, 1998, 2000), Hansen (2001) and Clements and Krolzig (1998), among many others. To evaluate point forecasts we use techniques based on loss functions which measure the amplitude of prediction error, so minimizing prediction error is the goal of these techniques.

Root Mean Squared Prediction Error:

Let $\left\{\Delta_h \hat{y}^{(i)}_{t|t-h}\right\}_{t=R+h}^{R+P}$ denote the sequence of the h-quarter growth rate $\Delta_h y_t$ of length $P_h = P-(h-1)$, obtained from model M_i. The corresponding forecast error is denoted $e^{(i)}_{t|t-h} = \Delta_h y_t - \Delta_h \hat{y}^{(i)}_{t|t-h}$. To evaluate point of forecasts, we consider the Root Mean Squared Prediction Error (RMSPE):

$$RMSPE=\sqrt{\frac{1}{p_h}\sum_{t=R+h}^{R+P}\left(\Delta_h y_t-\Delta_h\hat{y}^{(i)}_{t|t-h}\right)^2}$$

We compare also the RMSPE during recessions and expansions to evaluate the capacity of models to capture the business cycle asymmetry feature. We will date the business cycle regimes using the quarterly Bry-Boschan algorithm developed by Harding and Pagan (2002).

Diebold-Mariano test

The test of equal forecast accuracy developed by Diebold and Mariano (1995) is widely used to assess the significance of two competing models i and j. The DM is based on a loss differential noted d_t defined as the difference between the loss functions corresponding to each model, that is $d_t = g\left(e_{t|t-h}^{(i)}\right) - g\left(e_{t|t-h}^{(j)}\right)$, so the DM statistic is given by:

$$DM = \frac{\bar{d}}{\sqrt{V(\bar{d})}} \vec{d} N(0,1)$$

Where $\bar{d}$ is the sample mean loss differential such as $\bar{d} = \frac{1}{P_h}\sum_{t=R+h}^{R+P} d_t$, and where $\hat{V}(d) = \frac{1}{P_h}\left(\hat{\gamma}_0 + 2\sum_{k=0}^{h-1}\hat{\gamma}_k\right)$[1] is a consistent estimator of the asymptotic variance of $\bar{d}$ computed up to order $h-1$. As the test was found to be over-sized for the moderate samples, we apply the modification suggested by Harvey, Leybourne and Newbold (1997). The modified Diebold and Mariano statistic (M-DM) is then: $M-DM = \left(\frac{P_h + 1 - 2h + h(h-1)/P_h}{P_h}\right) DM$. The modified statistic is compared with critical values from the Student's t distribution with $(P_h - 1)$ degrees of freedom rather than the standard normal distribution[2].

Encompassing test

Forecast encompassing test evaluate whether a combined forecasts of competing models M_i and Mj, produce forecasts that are superior to the individual forecasts. So the combined forecasts will be given as follow:

$$\Delta_h \hat{y}_{t|t-h} = (1-\lambda)\Delta_h \hat{y}_{t|t-h}^{(i)} + \lambda \Delta_h \hat{y}_{t|t-h}^{(j)}$$

The model i is said to forecast encompass the model j if $\lambda = 0$. To compute the encompassing test, we follow Harvey, Leybourne and Newbold (1998) who suggested replacing the loss function by $d_t = e_{t|t-h}^{(i)}\left(e_{t|t-h}^{(i)}\ e_{t|t-h}^{(j)}\right)$ in the procedure of Diebold and Mariano (1995).

---

1 $$\hat{\gamma}_k = \frac{1}{P_h}\sum_{t=R+h+k}^{R+P}\left(d_t - \bar{d}\right)\left(d_{t-k} - \bar{d}\right)$$

2 We use the Newey and West (1987) procedure with the Bartlett Kernel and a truncation lag of h-1 in estimating the asymptotic variance used to compute the M-DM.

Interval forecast

The interval forecasts were evaluated in Christoffersen (1998) by means of the likelihood ratio tests of unconditional coverage, independence and conditional coverage. In this paper, the Pearson type χ^2 tests developed in wallis (2003) will be considered for two raisons that they asymptotically equivalent to the likelihood ratios tests, but their application seems to be simpler. The tests are based on indicator variables $\{I_{t|t-1}\}_{t=R+1}^{R+P}$ of length $P_1 = P$, it takes the value 1 when the realization ∆y_{t} lies inside the forecast interval and 0 otherwise. As given in wallis (2003), the three tests share the common form $X^2 = \sum(O-E)/E$, where the observed outcome (O) is compared to the expected outcome (E), see Siliverstovs and van Dijk (2003) for more details.

To extend these tests for h quarter growth rate, we follow the procedure of Diebold, Gunter and Tay (1998). This procedure partitionates the indicator function into h independent sub-groups. Then for each h sub-groups, the three tests χ_{UC}^2, χ_{IND}^2 and χ_{CC}^2 are performed and the null hypothesis is rejected for a given test at a significance level λ if the null hypothesis is rejected for any sub-group at λ/h significance level.

Density forecasts

In order to evaluate density forecasts, we use the method proposed in Diebold et al (1998). This method is based on comparing the true predictive density $f_t(.)$ and the 1-step ahead density forecast $p_{t|t-1}(.)$, using the probability integral transform (PIT) $z_{t|t-1}$, which is equal to the CDF corresponding to the density $p_{t|t-1}$. Under the null hypothesis, the PIT sequence is assumed to be $iidU\,[0;1]$ and according to Siliverstovs and van Dijk (2003), the uniformity assumption is assessed using the Kolmogorov-Smirnov (KS) test statistic. As the KS test is based on the assumption of independence, the Ljung-Box test must be performed. Berkowitz (2001) suggests to use the inverse normal CDF of the transformed PIT, and under the null hypothesis the PIT* is $iid\ N(0,1)$, again we follow Siliverstovs and VanDijk (2003) to test for the normality using the Doornik and

Hansen (1994) (DH). This procedure will be applied to h-quarter growth rate forecasts by partitioneting the sequence of PIT into h independent subgroups; see Diebold et al. (1998) for further details.

Empirical illustration

Data

Our data consist of quarterly seasonally unadjusted industrial production volume index for Tunisia. The data are taken from the Statistical National Institute and are transformed by taking logarithms. We consider the forecasting performance of the seven models discussed in the previous section over the period 1988:1 to 2006:2 using expanding window of data, starting with 1976:1-1987:4. The first window corresponds to an effective sample size of R=35, for each window we specify and we estimate all models, and we compute point, interval and density forecasts for each model of the h-quarter growth rates $\{\Delta_h yt_{+h} = y_{t+h} - y_t\}_{t=R}^{R+P-h}$ for $h = 1,....,12$, where $P = 74$. This procedure gives us $P_h = P - (h-1)$ forecasts for the h-quarter growth rate, $h = 1,....,12$.

Point forecasts evaluation

RMSPE

Table (1) contains the root mean squared prediction error (RMSPE) for the different models for horizon $h = 1,2,3,6,9$ and 12, which in general will be the horizons of most interest. Ranks of the different models are reported between parentheses.

The inspection of table (1) provides the following conclusions. First of all, the linear models perform well at all horizons, in particular linear models with simple description of seasonality (AR and SUR). The most notable improvement in forecasting performance as h increases is achieved by the TVD-AR model. Second, nonlinear models with a varying seasonal components, essentially The TVD-STAR and the TV-STAR models save the same bad forecast performance over the all horizon forecast. Finally, the SUR and the SUR-STAR models perform best.

Table (2) reports the RMSPE for the different models and the corresponding ranks, for the recessions and expansions separately. From the corresponding figure 1, we remark that the RMSPE of linear models are higher during recessions than expansions. This shows that linear models remain a useful tool to forecast accurately business cycle expansions.

Diebold-Mariano test

We evaluate the point forecast performance of the different models using the modified Diebold and Mariano test (HLN (1997)). Table 3 presents the M-DM statistic which rejects the null hypothesis that model M_i's and model M_j's forecasting performance at horizon h as measured by RMSPE are equal, in favor of one-sided alternative that M_i's RMSPE is superior at the 5% significance level. For example, if we consider the non linear time varying behavior, the high frequency of rejection of the M-DM statistic leads us to conclude that linear models (AR, SUR and TVD-AR) render significantly more accurate point forecasts than the TVD-STAR and TV-STAR models for all forecast horizons.

Encompassing test

Table 4 reports the encompassing statistic summarized across the all forecast horizons. The results of the pairwise forecast encompassing test confirm the observations made above: the SUR-STAR model dominates the other models both for the shorter and longer forecast horizons, in the sense that it tends to forecast encompass other models. The SUR model seems to be the second-best, the AR model is ranked the third.

Interval forecasts evaluation

Pearson χ^2 statistics used to evaluate interval forecasts for nominal coverage probabilities of 50%, 75% and 90% are provided in table 5 for $h = 1,2$ and 3. The TVD-STAR, TV-STAR and the SUR-STAR models offer the best performance across the three tests and the nominal coverage 50%, and the SUR and SUR-STAR models for the nominal coverage 75%. The test of correct unconditional coverage indicates that these models have empirical coverage rates that are closet to the nominal ones. All the models failed to achieve a good performance of the correct unconditional coverage and correct conditional coverage for nominal coverage 90%. The independence test is weakly rejected for all models across the nominal coverage considered indicating a good forecasting performance.Summarizing the results in table 5 concerning interval forecast evaluation, we find that the nonlinear models provide a good performance than linear model, and models with more elaborate seasonal components perform models with a simple description of seasonality.

Density forecasts evaluation

Table 6 provides results of the density forecast evaluation using the Kolmogrov-Smirnov and Doornik-Hansen tests, applied to the PIT and its inverse normal cumulative density function transformation. Looking at the KS statistic in column 4, there is some support for the nonlinear models (STAR and TV-STAR in particular) over linear models, as the KS statistic is significant at all reported horizons for the linear models, but it is significant at the 2-3 quarters for the TVD-STAR and the TV-STAR respectively, the worst density forecast is achieved by the SUR-STAR model.The DH statistic in column (5) is significant for all models at any reported horizons. Finally, we compute the Ljung-Box test for the first-order autocorrelation in the partitioned PIT sequences. The LB statistic is strongly rejected for even values of $k = 2,4$ unless for the STAR and SUR-STAR models at the 2 and 3 horizons forecast; respectively (for $k = 4$), LB statistic is only significant for the STAR model for odd values of $k = 1,3$ at the 2 and 3 quarters and for the SUR model at the 2 quarter.We conclude that according to the KS test results, non linear models seems to provide more accurate density forecast and then to reduce the uncertainty around the future observations. The DH tests indicate that having appeal to the normality test wasn't necessary and the LB for the first-order autocorrelation, demonstrate that all the models performs well (unless the SUR model).

Conclusion

In this paper, we have undertaken an extensive evaluation of the out-of-sample forecasting performance of a number of models for seasonality and nonlinearity, for the industrial production using quarterly data from 1976:2 to 2006:2. The models allow for a simple description of seasonality, also they take in account the smoothly changing (and nonlinear) deterministic seasonality. The forecasting performance of these models have been assessed, first, on their ability to produce point forecasts measured by means of RMSFE and the modified Diebold and Mariano tests. Second, on their capacity to generate interval and density forecasts evaluated using the procedures of Wallis (2003) and Diebold et al.(1998), respectively. Our findings can be summarized as follows: the results of the point forecast evaluation seem to be in line with previous findings in the literature, in that the linear models (AR and SUR) offer more accurate forecasts. They provide point forecasts that are often found to be superior to those of the smoothly changing and nonlinear models, across different forecast horizons. Especially, the SUR model performs well among the linear models; it is always ranked the second after the SUR-STAR model, which is the best performing nonlinear model. The time varying nonlinear seasonality: TVD-STAR, TV-STAR and SUR-STAR models produce superior interval forecasts when compared to those from the other models, this result holds in particular at the nominal coverage 50%. The same models have described well the business cycle feature. Finally, on the basis of density forecast evaluation tests the STAR and the TV-STAR models are the best performers as well. The obvious extension is to consider the multivariate case with combined forecasts.

References

Berkowitz J (2001) Testing density forecasts, with applications to risk management. Journal of Business and Economic Statistics 19: 465-474.
Christoffersen PF (1998) Evaluating interval forecasts. International Economic Review 39: 841-862.
Clements MP, Krolzig HM(1998) A comparison of the forecast performance of Markov-Switching and threshold autoregressive models of US GNP. Econometrics Journal 1: C47-C75.
De Gooijer JG, Kumar K (1992) Some recent developments in non-linear time series modelling, testing and forecasting, International Journal of Forecasting 8: 135-156.
Diebold FX., Mariano RS (1995) Comparing predictive accuracy. Journal of Business and Economic Statistics 13: 253-263.
Diebold FX., Gunther TA, Tay AS (1998) Evaluating density forecasts with applications to financial risk management. International Economic Review 39: 863--883.
Franses PH, van Dijk D (2005) Forecasting performance of various models for seasonality and nonlinearity for quarterly industrial production. International journal of forecasting 21: 87-102.
Fuller WA. (1976) Introduction to Statistical Time Series. Wiley, New York.
Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series subject to changes in regime. Econometrica 57: 357-384.
Harding D, Pagan AR (2002) Dissecting the cycle: a methodological investigation. Journal of Monetary Economics 49: 365--381.
Harvey DI, Leybourne SJ, P Newbold P (1997) Testing the equality of prediction mean squared errors. International Journal of Forecasting 13: 281-291.
Harvey DI, Leybourne SJ, P Newbold P (1998) Tests for forecast encompassing. Journal of Business and Economic Statistics 16: 254-259.

Hylleberg S, Engle RF, Granger CW, Yoo BS (1990) Seasonal integration and cointegration. Journal of Econometrics 44: 215-238.

Lundbergh S, Teräsvirta T, van Dijk D (2003) Time-varying smooth transition autoregressive models. Journal of Busines and Economic Statistics 21: 104--121.

Marcellino M (2004) Forecasting EMU macroeconomic variables. International Journal of Forecasting 20:359-372.

Medeiros MC, A. Veiga, CE Pedreira (2001) Modeling Exchange Rates: Smooth Transitions, Neural Networks, and Linear Models. IEEE Transactions on Neural Networks 12: 755-764.

Rapach D.E, Wohar ME (2005) The out-of-sample forecasting performance of nonlinear models of real exchange rate behavior. International Journal of Forecasting 22: 341-361.

Siliverstovs BD, V Dijk (2003) Forecasting Industrial Production with Linear, Nonlinear and Structural Change Models. Econometric Institute Report EI 2003-16, No. 16, May, 2003.

Stock JH, Watson MW (1998) A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, mimeo, Harvard University.

Teräsvirta T (1994) Specification, estimation, and evaluation of smooth transition autoregressive models. Journal of the American Statistical Association 89: 208-218.

Teräsvirta T, Anderson, HM (1992) Characterizing nonlinearities in business cycles using smooth transition autoregressive models. Journal of Applied Econometrics 7: S119--S136.

Tong H(1990) nonlinear time series: A dynamical systems approach. Oxford: Oxford University Press.

van Dijk D,Strikholm B, Teräsvirta T (2003) The effects of institutional and technological change and business cycle fluctuations on seasonal patterns in quarterly industrial production series. Econometrics Journal 6: 79-- 98.

Wallis KF (2003) Chi-squired tests of interval and density forecasts and the Bank of England's fan charts. International Journal of Forecasting 19: 165-175.

Appendix

The RMESPE during recessions and expansions

Fig-1

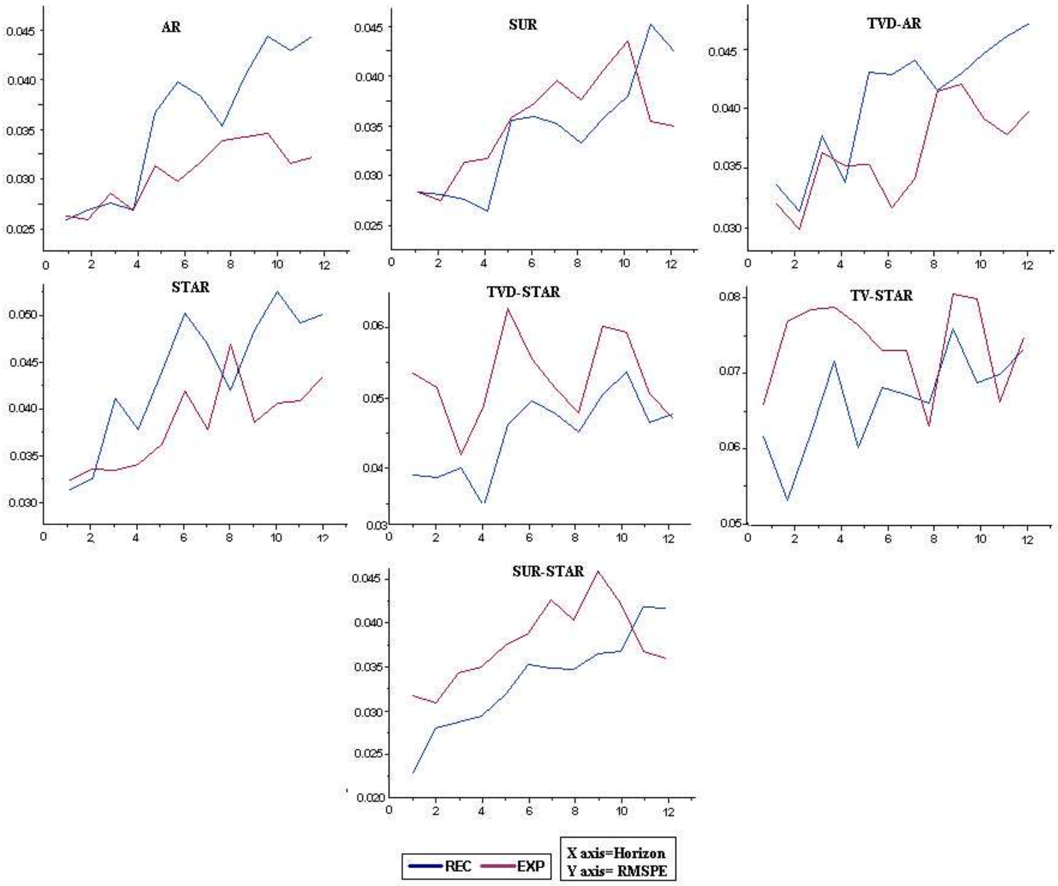

Table-1 Out-of-sample forecasts evaluation: The RMSPE

h	AR	SUR	TVD-AR	STAR	TVD-STAR	TV-STAR	SUR-STAR	Average
1	0.0315	0.0309	0.0327	0.032	0.0492	0.0643	0.0295	0,0386
	(3)	(2)	(5)	(4)	(6)	(7)	(1)	
2	0.0317	0.0303	0.0307	0.0331	0.047	0.0689	0.0302	0,0388
	(4)	(2)	(3)	(5)	(6)	(7)	(1)	
3	0.0336	0.0325	0.0371	0.0367	0.041	0.072	0.0324	0,0404
	(3)	(2)	(4)	(5)	(6)	(7)	(1)	
6	0.0399	0.0393	0.0374	0.0458	0.0527	0.0706	0.0375	0,0462
	(4)	(2)	(3)	(5)	(6)	(7)	(1)	
9	0.0428	0.0407	0.0427	0.044	0.055	0.0779	0.0414	0.0539
	(4)	(2)	(3)	(5)	(6)	(7)	(1)	
12	0.0443	0.0418	0.0442	0.047	0.0473	0.0737	0.0395	0.0482
	(4)	(2)	(3)	(5)	(6)	(7)	(1)	

Table-2 Out-of-sample point forecasts evaluation during recessions and expansions

h	AR	SUR	TVD-AR	STAR	TVD-STAR	TV-STAR	SUR-STAR
Recessions							
1	0.0313	0.0310	0.0338	0.0313	0.0389	0.0614	0.0232
	(3)	(2)	(5)	(4)	(6)	(7)	(1)
2	0.0323	0.0307	0.0316	0.0324	0.0386	0.0529	0.0283
	(4)	(2)	(3)	(5)	(6)	(7)	(1)
3	0.0330	0.0302	0.0379	0.0410	0.0399	0.0621	0.0290
	(3)	(2)	(4)	(6)	(5)	(7)	(1)
6	0.0449	0.0386	0.0430	0.0501	0.0494	0.0680	0.0355
	(4)	(2)	(3)	(6)	(5)	(7)	(1)
9	0.0455	0.0384	0.0432	0.0483	0.0503	0.0758	0.0368
	(4)	(2)	(3)	(5)	(6)	(7)	(1)
12	0.0493	0.0451	0.0474	0.0500	0.0476	0.0730	0.0420
	(5)	(2)	(3)	(6)	(4)	(7)	(1)
Expansions							
1	0.0317	0.0309	0.0322	0.0323	0.0533	0.0657	0.0320
	(2)	(1)	(3)	(4)	(6)	(7)	(3)
2	0.0313	0.0301	0.0301	0.0335	0.0512	0.0766	0.0312
	(4)	(2)	(1)	(5)	(6)	(7)	(3)
3	0.0339	0.0340	0.0365	0.0333	0.0418	0.0783	0.0346
	(2)	(3)	(5)	(1)	(6)	(7)	(4)
6	0.0351	0.0399	0.0319	0.0417	0.0553	0.0728	0.0391
	(2)	(4)	(1)	(5)	(6)	(7)	(3)
9	0.0394	0.0432	0.0422	0.0384	0.0599	0.0804	0.0462
	(2)	(4)	(3)	(1)	(6)	(7)	(5)
12	0.0375	0.0376	0.0401	0.0433	0.0469	0.0746	0.0363
	(2)	(3)	(4)	(5)	(6)	(7)	(1)

Note: The table contains the RMSPE of the different models and the corresponding ranks during business cycle recessions and expansions.

Table-3 Out-of-sample point forecast evaluation: The M-DM statistic

j \ i	AR		SUR		TVD-AR		STAR		TVD-STAR		TV-STAR		SUR-STAR	
h=1														
AR			-0.33	[0.63]	0.65	[0.26]	0.35	[0.36]	1.95	**[0.03]**	3.27	**[0.00]**	-0.91	[0.82]
SUR	0.33	[0.37]			0.68	[0.25]	0.56	[0.29]	2.02	**[0.02]**	3.25	**[0.00]**	-1.05	[0.85]
TVD-AR	-0.65	[0.74]	-0.68	|0.75]			-0.38	[0.65]	1.87	**[0.03]**	3.22	**[0.00]**	-1.05	[0.85]
STAR	-0.35	[0.64]	-0.56	|0.71]	0.38	[0.35]			1.90	**[0.03]**	3.28	**[0.00]**	-1.24	[0.89]
TVD-STAR	-1.95	[0.97]	-2.02	[0.98]	-1.87	[0.97]	-1.90	[0.97]			1.42	[0.08]	-2.15	[0.98]
TV-STAR	-3.27	[1.00]	-3.25	[1.00]	-3.22	[1.00]	-3.28	[1.00]	-1.42	[0.92]			-3.32	[1.00]
SUR-STAR	0.91	[0.18]	1.05	[0.15]	1.05	[0.15]	1.24	[0.11]	2.15	**[0.02]**	3.32	**[0.00]**		
h=2														
AR			-0.45	[0.67]	-0.42	[0.66]	0.83	[0.21]	1.38	[0.09]	2.57	**[0.01]**	-0.44	[0.67]
SUR	0.45	[0.33]			0.08	[0.47]	0.83	[0.21]	1.50	[0.07]	2.59	**[0.01]**	-0.07	[0.53]
TVD-AR	0.42	[0.34]	-0.08	[0.53]			0.87	[0.19]	1.47	[0.07]	2.56	**[0.01]**	-0.11	[0.54]
STAR	-0.83	[0.79]	-0.83	[0.79]	-0.87	[0.81]			1.25	[0.11]	2.50	**[0.01]**	-0.81	[0.79]
TVD-STAR	-1.38	[0.91]	-1.50	[0.93]	-1.47	[0.93]	-1.25	[0.89]			1.53	[0.07]	-1.55	[0.94]
TV-STAR	-2.57	[0.99]	-2.59	[0.99]	-2.56	[0.99]	-2.50	[0.99]	-1.53	[0.93]			-2.60	[0.99]
SUR-STAR	0.44	[0.33]	0.07	[0.47]	0.11	[0.46]	0.81	[0.21]	1.55	[0.06]	2.60	**[0.01]**		
h=3														
AR			-0.15	[0.56]	1.06	[0.15]	1.08	[0.14]	1.60	[0.06]	2.52	**[0.01]**	-0.14	[0.56]
SUR	0.15	[0.44]			1.18	[0.12]	0.73	[0.23]	1.77	**[0.04]**	2.50	**[0.01]**	0.00	[0.50]
TVD-AR	-1.06	[0.85]	-1.18	[0.88]			-0.21	[0.58]	0.75	[0.23]	2.23	**[0.01]**	-1.35	[0.91]
STAR	-1.08	[0.86]	-0.73	[0.77]	0.21	[0.42]			0.88	[0.19]	2.36	**[0.01]**	-0.72	[0.76]
TVD-STAR	-1.60	[0.94]	-1.77	[0.96]	-0.75	[0.77]	-0.88	[0.81]			2.04	**[0.02]**	-1.86	[0.97]
TV-STAR	-2.52	[0.99]	-2.50	[0.99]	-2.23	[0.99]	-2.36	[0.99]	-2.04	[0.98]			-2.50	[0.99]
SUR-STAR	0.14	[0.44]	0.00	[0.50]	1.35	[0.09]	0.72	[0.24]	1.86	**[0.03]**	2.50	**[0.01]**		
h=6														
AR			0.01	[0.50]	-0.74	[0.77]	1.03	[0.15]	1.33	[0.09]	2.25	**[0.01]**	-0.32	[0.63]
SUR	-0.01	[0.50]			-0.30	[0.62]	0.72	[0.24]	1.36	[0.09]	2.24	**[0.01]**	-1.10	[0.86]
TVD-AR	0.74	[0.23]	0.30	[0.38]			1.11	[0.14]	1.45	[0.08]	2.38	**[0.01]**	0.03	[0.49]
STAR	-1.03	[0.85]	-0.72	[0.76]	-1.11	[0.86]			0.69	[0.25]	1.94	**[0.03]**	-0.94	[0.82]
TVD-STAR	-1.33	[0.91]	-1.36	[0.91]	-1.45	[0.92]	-0.69	[0.75]			1.32	[0.10]	-1.57	[0.94]
TV-STAR	-2.25	[0.99]	-2.24	[0.99]	-2.38	[0.99]	-1.94	[0.97]	-1.32	[0.90]			-2.36	[0.99]
SUR-STAR	0.32	[0.37]	1.10	[0.14]	-0.03	[0.51]	0.94	[0.18]	1.57	[0.06]	2.36	**[0.01]**		
h=9														
AR			-0.19	[0.57]	0.41	[0.34]	0.48	[0.32]	1.41	[0.08]	2.01	**[0.02]**	-0.06	[0.52]
SUR	0.19	[0.43]			0.44	[0.33]	0.33	[0.37]	1.66	[0.05]	2.30	**[0.01]**	0.22	[0.42]
TVD-AR	-0.41	[0.66]	-0.44	[0.67]			-0.06	[0.52]	1.26	[0.11]	2.01	**[0.02]**	-0.26	[0.60]
STAR	-0.48	[0.68]	-0.33	[0.63]	0.06	[0.48]			1.31	[0.10]	1.89	**[0.03]**	-0.18	[0.57]
TVD-STAR	-1.41	[0.92]	-1.66	[0.95]	-1.26	[0.89]	-1.31	[0.90]			1.09	[0.14]	-1.47	[0.93]
TV-STAR	-2.01	[0.98]	-2.30	[0.99]	-2.01	[0.98]	-1.89	[0.97]	-1.09	[0.86]			-2.21	[0.98]
SUR-STAR	0.06	[0.48]	-0.22	[0.58]	0.26	[0.40]	0.18	[0.43]	1.47	[0.07]	2.21	**[0.02]**		
h=12														
AR			-0.15	[0.56]	0.70	[0.24]	0.91	[0.18]	0.63	[0.26]	2.22	**[0.02]**	-0.55	[0.71]
SUR	0.15	[0.44]			0.51	[0.31]	0.67	[0.25]	0.92	[0.18]	2.53	**[0.01]**	-1.45	[0.92]
TVD-AR	-0.70	[0.76]	-0.51	[0.69]			0.32	[0.37]	0.40	[0.34]	2.13	**[0.02]**	-1.04	[0.85]
STAR	-0.91	[0.82]	-0.67	[0.75]	-0.32	[0.63]			0.23	[0.41]	1.99	**[0.03]**	-1.17	[0.88]
TVD-STAR	-0.63	[0.74]	-0.92	[0.82]	-0.40	[0.66]	-0.23	[0.59]			2.29	**[0.01]**	-1.42	[0.92]
TV-STAR	-2.22	[0.98]	-2.53	[0.99]	-2.13	[0.98]	-1.99	[0.97]	-2.29	[0.99]			-2.73	[1.00]
SUR-STAR	0.55	[0.29]	1.45	[0.08]	1.04	[0.15]	1.17	[0.12]	1.42	[0.08]	2.73	**[0.00]**		

Note : Table entries are the modified Diebold-Mariano (1995) test statistic for the null hypothesis that the column model i RMSPE equals the model row j, p-value using student's t distribution with P_h-1 degrees of freedom is reported in brackets, bold p-value indicates significance at 5%.

Table-4 Pair wise comparison: encompassing test summary

j \ i	AR	SUR	TVD-AR	STAR	TVD-STAR	TV-STAR	SUR-STAR	Total
h = 1 - 6								
AR		4	2	1	0	1	4	12
SUR	3		1	3	2	0	1	10
TVD-AR	3	4		1	0	0	5	13
STAR	4	6	3		1	1	6	21
TVD-STAR	3	5	2	3		0	4	17
TV-STAR	6	6	6	6	6		6	36
SUR-STAR	1	1	0	1	0	0		3
h = 7 - 12								
AR		5	2	0	5	0	4	16
SUR	0		0	0	1	0	0	1
TVD-AR	2	4		1	4	0	6	17
STAR	1	6	2		4	1	6	20
TVD-STAR	0	6	0	0		0	6	12
TV-STAR	6	6	6	6	6		6	36
SUR-STAR	0	0	1	0	2	0		3
h = 1 - 12								
AR		9	4	1	5	1	8	28
SUR	3		1	3	3	0	1	11
TVD-AR	5	8		2	4	0	11	30
STAR	5	12	5		5	2	12	41
TVD-STAR	3	11	2	3		0	10	29
TV-STAR	12	12	12	12	12		12	72
SUR-STAR	1	1	1	1	2	0		6

Note : The *(i,j)* entries in the table are the rejections of the null hypothesis that model *i's* forecast encompasses model *j's* forecast by the encompassing test at the 5% significance level.

Table-5 Interval forecast evaluation for nominal coverage probabilities: 50%, 75% and 90%

(1)	(2)	(3)	(4)		
			Nominal Coverage 50%		
models	h^a	0,10/h	χ^2_{UC} [b]	χ^2_{IND} [c]	χ^2_{CC} [d]
AR	1	0.10	**5.41** [0.02]	0.14 [0.80]	**6.17** [0.04]
	2	0.05	**1.32** [0.26], 3.27 [0.07]	**6.22** [0.02], **6.69**[0,02]	**7.69** [0.02], **7.69** [0.02]
	3	0.033	**4.84** [0.04],**17.64** [0.00],**10.67** [0.00]	0.69 [1.00], 0.69 [1.00],1.25 [0.36]	**11.05** [0.00],**11.05** [0.00],**13.13** [0.00]
SUR	1	0.10	0.27 [0.73]	2.28 [0.16]	2.40 [0.32]
	2	0.05	**6.08** [0.01], **9.76** [0.00]	4.62 [0.08], 4.62 [0.08]	**10.81** [0.00],**10.81** [0.00]
	3	0.033	**17.64** [0.00],**11.56** [0.00], **8.17** [0.02]	**5.87** [0.04], **5.87** [0.04],**8.07** [0.02]	**12.04** [0.00],**12.04** [0.00],**12.84** [0.00]
TVD-AR	1	0.10	3.46 [0.08]	0.39 [0.62]	4.33 [0.11]
	2	0.05	2.19 [0.19], 0.24 [0.63]	1.74 [0.32], 1.74 [0.32]	1.85 [0.43], 1.85 [0.43]
	3	0.033	**9.00** [0.00], **6.76** [0.01], **6.00** [0.02]	**10.19** [0.01],**10.19** [0.01],**8.07** [0.02]	**13.64** [0.00],**13.64** [0.00],**12.84** [0.00]
STAR	1	0.10	**51.95** [0.00]	0.59 [1.00]	**51.15** [0.00]
	2	0.05	**19.70** [0.00],**14.30** [0.00]	0.30 [1.00], 0.30 [1.00]	**25.10** [0.00],**25.10** [0.00]
	3	0.033	**11.56** [0.00], **4.84** [0.04], **8.17** [0.00]	1.41 [0.54], 1.41 [0.54],2.27 [0.19]	**9.09** [0.01], **9.09** [0.01], **8.90** [0.01]
TVD-STAR	1	0.10	**17.51** [0.00]	0.66 [0.54]	**17.29** [0.00]
	2	0.05	2.19 [0.19], 0.03 [1.00]	3.41 [0.10], 3.41 [0.10]	4.69 [0.07], 4,69 [0.07]
	3	0.033	0.04 [1.00], 0.04 [1.00], 2.67 [0.15]	2.67 [0.22], 2.67 [0.22],3.57 [0.10]	2.67 [0.33], 2.67 [0.33], 3.61 [0.19]
TV-STAR	1	0.10	2.45 [0.36]	2.37 [0.21]	3.30 [0.11]
	2	0.05	3.27 [0.10], 3.38 [0.09]	0.00 [1.00], 0.00 [1.00]	3.46 [0.08], 3.46 [0.08]
	3	0.033	1.00 [0.42], 0.04 [1.00], 0.67 [0.54]	0.00 [1.00], 0.00 [1.00],0.01 [1.00]	**8.17** [0.01], **8.17** [0.01], **7.36** [0.02]
SUR-STAR	1	0.10	4.38 [0.05]	0.39 [0.62]	4.33 [0.11]
	2	0.05	4.57 [0.05], 4.57 [0.05]	0.00 [1.00], 0.00 [1.00]	4.66 [0.06], 4.66 [0.06]
	3	0.033	4.39 [0.07], 2.48 [0.18], 1.44 [0.31]	1.36 [0.34], 1.36 [0.34],2.64 [0.25]	2.81 [0.10], 3.86 [0.08], 3.45 [0.13]

(1)	(2)	(3)	(4)		
			Nominal Coverage 75%		
models	h^a	0,10/h	χ^2_{UC} [b]	χ^2_{IND} [c]	χ^2_{CC} [d]
AR	1	0.10	**58.54** [0.00]	3.35 [0.08]	**64.49** [0.00]
	2	0.05	**4.77** [0.04],**16.66** [0.00]	0.01 [1.00],0.01 [1.00]	**7.27** [0.03], **7.27**[0.03]
	3	0.033	0.12 [0.82], 0.65 [0.49], 0.22 [0.81]	2.52 [0.17],2.52 [0.17],3.39 [0.14]	3.88 [0.15], 3.88 [0.15], 4.19 [0.12]
SUR	1	0.10	**13.14** [0.00]	4.43 [0.06]	**19.63** [0.00]
	2	0.05	0.23 [0.71], 1.52 [0.26]	0.15 [1.00],1.15 [1.00]	0.72 [0.71], 0.72 [0.71]
	3	0.033	2.25 [0.17], 0.33 [0.65], 2.00 [0.24]	0.96 [0.57],0.96 [0.57],1.02 [0.56]	1.60 [0.55], 1.60 [0.55], 1.49 [0.55]
TVD-AR	1	0.10	**54.50** [0.00]	1.25 [0.33]	**53.86** [0.00]
	2	0.05	**8.66** [0.00],**11.04** [0.00]	**7.06** [0.02],**7.06** [0.02]	**18.86** [0.00],**18.86** [0.00]
	3	0.033	0.65 [0.49], 0.65 [0.49], 0.89 [0.48]	1.68 [0.31],1.68 [0.31],2.41 [0.28]	2.07 [0.53], 2.07 [0.53], 2.49 [0.32]
STAR	1	0.10	**176.59** [0.00]	0.62 [0.41]	**173.88** [0.00]
	2	0.05	**74.60** [0.00],**50.68** [0.00]	1.44 [0.56],1.44 [0.56]	**66.40** [0.00],**66.40** [0.00]
	3	0.033	**24.65** [0.00],**29.45** [0.00],**22.22** [0.00]	2.48 [0.18],2.48 [0.18],0.61 [1.00]	**39.73** [0.00],**39.73** [0.00],**41.18** [0.00]
TVD-STAR	1	0.10	**90.83** [0.00]	0.09 [0.77]	**88.32** [0.00]
	2	0.05	**23.43** [0.00],**23.43** [0.00]	1.44 [0.31],1.44 [0.31]	**23.20** [0.00],**23.20** [0.00]
	3	0.033	**7.05** [0.01],**16.33** [0.00],**5.56** [0.03]	0.62 [0.68],0.62 [0.68],0.38 [0.68]	**11.71** [0.00],**11.71** [0.00], **9.57** [0.01]
TV-STAR	1	0.10	**80.88** [0.00]	5.90 [0.03]	**84.99** [0.00]
	2	0.05	**23.43** [0.00],**27.25** [0.00]	5.06 [0.06],5.06 [0.06]	**39.33** [0.00],**39.33** [0.00]
	3	0.033	**16.33** [0.00], 3.00 [0.10],**14.22** [0.00]	4.59 [0.06],4.59 [0.06],4.15 [0.07]	**27.67** [0.00],**27.67** [0.00],**24.87** [0.00]
SUR-STAR	1	0.10	**15.15** [0.00]	2.81 [0.10]	**19.60** [0.00]
	2	0.05	0.44 [0.57], 1.52 [0.26]	0.86 [0.65],0.86 [0.65]	0.86 [0.69], 0.86 [0.69]
	3	0.033	2.25 [0.17], 0.33 [0.65], 2.00 [0.24]	0.96 [0.57],0.96 [0.57],1.02 [0.56]	1.60 [0.55], 1.60 [0.55], 1.49 [0.55]

(1)	(2)	(3)	(4)		
			Nominal Coverage 90%		
models	h[a]	0,10/h	χ^2_{UC}[b]	χ^2_{IND}[c]	χ^2_{CC}[d]
AR	1	0.10	**272.49** [0.00]	1.89 [0.19]	**282.04** [0.00]
	2	0.05	**45.43** [0.00], **89.88** [0.00]	2.03 [0.19],2.03 [0.19]	**53.03** [0.00], **53.03** [0.00]
	3	0.033	**32.11** [0.00], **13.44** [0.00], **26.74** [0.00]	0.41 [0.68],0.41 [0.68],0.21 [0.69]	**27.85** [0.00], **27.85** [0.00], **22.24** [0.00]
SUR	1	0.10	**131.56** [0.00]	4.95 [0.04]	**148.02** [0.00]
	2	0.05	**16.00** [0.00], **8.44** [0.01]	0.25 [0.70],0.25 [0.70]	**17.49** [0.00], **17.49** [0.00]
	3	0.033	**5.44** [0.03], 1.00 [0.24], 1.19 [0.21]	0.69 [1.00],0.69 [1.00],0.52 [1.00]	2.24 [0.39], 2.24 [0.39], 0.89 [0.60]
TVD-AR	1	0.10	**285.43** [0.00]	2.84 [0.11]	**297.31** [0.00]
	2	0.05	**70.30** [0.00],**136.24** [0.00]	1.63 [0.31],1.63 [0.31]	**87.47** [0.00], **87.47** [0.00]
	3	0.033	**40.11** [0.00], **32.11** [0.00], **34.24** [0.00]	1.14 [0.40],1.14 [0.40],1.67 [0.38]	**29.83** [0.00], **29.83** [0.00], **26.12** [0.00]
STAR	1	0.10	**588.40** [0.00]	0.25 [1.00]	**579.58** [0.00]
	2	0.05	**275.70** [0.00],**207.71** [0.00]	0.30 [1.00],0.30 [1.00]	**267.03** [0.00],**267.03** [0.00]
	3	0.033	**136.11** [0.00], **81.00**[0.00],**112.67** [0.00]	1.36 [0.34],1.36 [0.34],1.25 [0.36]	**161.82** [0.00],**161.82** [0.00],**152.93** [0.00]
TVD-STAR	1	0.10	**415.43** [0.00]	0.99 [0.45]	**408.53** [0.00]
	2	0.05	**111.86** [0.00],**111.86** [0.00]	1.44 [0.31],1.44 [0.31]	**97.33** [0.00], **97.33** [0.00]
	3	0.033	**58.78** [0.00], **49.00** [0.00], **42.67** [0.00]	0.67 [0.68],0.67 [0.68],0.38 [0.68]	**44.52** [0.00], **44.52** [0.00], **37.62** [0.00]
TV-STAR	1	0.10	**447.62** [0.00]	0.00 [1.00]	**439.92** [0.00]
	2	0.05	**111.86** [0.00],**177.32** [0.00]	0.20 [1.00],0.20 [1.00]	**184.14** [0.00],**184.14** [0.00]
	3	0.033	**49.00** [0.00], **58.78** [0.00], **34.24** [0.00]	0.00 [1.00],0.00 [1.00],0.02 [1.00]	**98.69** [0.00], **98.69** [0.00], **90.71** [0.00]
SUR-STAR	1	0.10	**106.24** [0.00]	1.54 [0.25]	**112.76** [0.00]
	2	0.05	**31.86** [0.00], **11.92** [0.00]	0.45 [0.72],0.45 [0.72]	**34.58** [0.00], **34.58** [0.00]
	3	0.033	**5.44** [0.03], 2.78 [0.10], 3.13 [0.09]	1.26 [0.54],1.26 [0.54],1.02 [0.56]	5.44 [0.09], 5.44 [0.09], 3.02 [0.20]

Notes: Statistics are reported for each of the h subgroups; the exact *p*-value is reported in brackets; bold statistic indicate significance at the 0.10 /*h* level according to the exact *p*-value.

[a] Forecast horizon (in months).

[b] Pearson χ^2 test statistic for the null hypothesis that the prediction intervals have correct unconditional coverage.

[c] Pearson χ^2 test statistic for the null hypothesis that the prediction intervals are independent.

[d] Pearson χ^2 test statistic for the null hypothesis that the prediction intervals have correct conditional coverage.

Table-5 Density forecast evaluation

(1) models	(2) h[a]	(3) *0.10/h*	(4) KS[b]	(5) DH[c]	(6) LB, k=1[d]	(7) LB, k=2[d]	(8) LB, k=3[d]	(9) LB, k=4[d]
AR	1	0.10	**0.10**	**981.69**	0.73	**38.10**	0.51	**25.35**
	2	0.05	**0.15**, **0.16**	**1.55**, **188.31**	1.48, 0.33	**11.85**, **16.64**	1.31, 0.00	3.26, **12.90**
	3	0.033	**0.26**, **0.13**, **0.14**	**120.82**, **164.76**, **151.16**	0.79, 0.00, 0.88	**7.81**, **7.62**, **12.39**	0.01, 0.27, 0.30	1.87, **4.42**, **9.36**
SUR	1	0.10	**0.07**	**1672.14**	0.00	**35.13**	0.05	**23.08**
	2	0.05	**0.12**, **0.11**	2.30, **399.41**	1.50, 0.12	**14.70**, **9.66**	**4.91**, 0.33	**9.41**, 3.46
	3	0.033	**0.17**, **0,17**, **0.10**	**161.06**, **371.76**, 2.97	1.50, 1.10, 2.05	**6.65**, 2.15, **9.68**	1.14, 0.31, 1.63	1.42, 0.29, **5.90**
TVD-AR	1	0.10	**0.11**	**608.57**	0.48	**42.55**	0.67	**32.76**
	2	0.05	**0.15**, **0.14**	**226.59**, **189.46**	0.07, 0.49	**15.04**, **16.96**	0.29, 0.00	**10.78**, **8.30**
	3	0.033	**0.16**, **0.16**, **0.15**	**99.36**, **126.23**, **13,27**	0.01, 0.17, 0.46	**5.84**, **10.76**, **12.17**	0.01, 0.08, 0.47	2.53, **4.89**, **8.01**
STAR	1	0.10	0.36	**81.71**	**30.15**	**58.29**	**35.20**	**51.51**
	2	0.05	0.64, 0.74	**64.81**, 4.60	0.13, 1.92	0.58, 0.21	0.00, 0.00	0.00, 0.00
	3	0.033	0.42, 0.36, 0.40	**43.76**, **49.53**, **42.80**	**17.04**, **11.86**, **18.72**	**19.20**, **11.39**, **19.94**	**17.86**, **5.50**, **17.94**	**17.99**, 2.95, **17.77**
TVD-STAR	1	0.10	0.22	**228.10**	0.16	**46.08**	0.00	**40.46**
	2	0.05	0.25, **0.17**	**104.17**, **106.44**	0.77, 0.26	**23.10**, **17.20**	1.24, 0.01	**17.55**, **10.69**
	3	0.033	**0.24**, **0.21**, **0.27**	**67.89**, **77.98**, **78.43**	1.60, 0.06, 0.17	**11.08**, **7.38**, **11.45**	1.35, 0.12, 0.13	**7.71**, 3.80, **4.50**
TV-STAR	1	0.10	0.30	**135.96**	0.01	**45.88**	0.00	**24.37**
	2	0.05	0.29, 0.23	**65.10**, **73.02**	0.08, 0.43	**21.43**, **22.36**	0.20, 1.50	**15.57**, 14.48
	3	0.033	0.37, 0.38, **0.26**	**34.01**, **37.64**, **55.85**	0.30, 0.06, 0.02	**21.79**, **9.68**, **11.61**	0.33, 1.58, 1.00	**15.73**, **4.66**, **6.21**
SUR-STAR	1	0.10	**0.10**	**1372.92**	0.02	**19.04**	2.19	**9.73**
	2	0.05	**0.14**, **0.13**	2.34, **376.25**	1.47, 0.10	**9.38**, **13.03**	2.54, 0.26	2.94, **5.04**
	3	0.033	**0.16**, **0.19**, **0.19**	**161.45**, **235.55**, 0.76	1.75, 0.04, 0.85	**5.43**, 1.89, **7.27**	1.68, 0.01, 0.41	1.56, 0.15, 3.25

Notes: Statistics are reported for each of the *h* subgroups; bold statistic indicates significance at the 0.10 /*h* level.
[a] Forecast horizon (in months).
[b] Kolmogorov–Smirnov test statistic for the null hypothesis that $z_t \sim U(0, 1)$.
[c] Doornik and Hansen (1994) test statistic for the null hypothesis that $z_t^* \sim N(0, 1)$.
[d] Ljung–Box test statistic for the null hypothesis of no first-order autocorrelation in $(z_t - \bar{z})^k$, k =1, . . ., 4.